DOCUMENT RESUME

ED 144 957 TM 006 389

AUTHOR Warshavsky, Belle
TITLE The Concurrent Validity of the Third Grade New York State Pupil Evaluation Program in Reading with Standardized Tests of Reading, Intelligence, and Cloze.
PUB DATE Jan 75
NOTE 112p.; Ph.D. Dissertation, Walden University ; Parts of appendix may be marginally legible

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Cloze Procedure; Elementary School Students; Grade 3; *Intelligence Tests; *Multiple Regression Analysis; Primary Education; Reading Comprehension; *Reading Tests; Sex Differences; Sight Vocabulary; Standardized Tests; State Programs; Testing Programs; *Test Validity
IDENTIFIERS Durrell Listening Reading Series; Gilmore Oral Reading Test; *New York State Pupil Evaluation Program; Wechsler Intelligence Scale for Children; Wide Range Achievement Test

ABSTRACT

The concurrent validity of the third grade, New York State Pupil Evaluation Reading Program (PEP) reading test was examined with the following standardized tests: the Durrell Silent Reading Test, the Gilmore Oral Reading Test, the Wide Range Achievement Test (WRAT), the Verbal IQ of the Wechsler Intelligence Scale for Children (WISC), and the Cloze Test, constructed by the investigator. Fifty third graders in a Long Island, New York elementary school were selected for the study. Multiple regression analyses revealed that the variables that accounted for the highest correlation with the PEP test were the WRAT, the Gilmore Oral Reading Comprehension sub-test, and the Similarities sub-test of the WISC. Sex differences were not found in any measure of the study. Similarly, verbal IQ did not add anything significant to the variance. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144957

THE CONCURRENT VALIDITY OF THE THIRD GRADE NEW YORK STATE PUPIL EVALUATION PROGRAM IN READING WITH STANDARDIZED TESTS OF READING, INTELLIGENCE, AND CLOZE

BEST COPY AVAILABLE

By

Belle Warshavsky

B.B.A. St. John's University
M.S.Ed. Hofstra University
Professional Diploma Hofstra University

Raymond Joseph Horton, Ph.D., Advisor
Reading Consultant--Plainview-Old Bethpage
Public Schools, Plainview, New York

BEST COPY AVAILABLE

A Dissertation Submitted in Partial Fulfillment of
The Requirements for the Degree of
Doctor of Philosophy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Belle Warshavsky

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Walden University
January 1975

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM006 389

ACKNOWLEDGEMENTS

This investigator wishes to thank a number of persons who contributed their time and efforts towards this investigation.

First and foremost, I would like to thank my husband, Henry, without whose encouragement, patience, and assistance, this investigation would not have been completed.

I would also like to give special thanks to my advisor, Dr. Raymond J. Horton, whose professional knowledge in the area of statistics proved invaluable.

Lastly, I would like to thank Dr. Robert F. Savitt, Superintendent of Schools, who permitted me to do the extensive testing needed for this investigation during a district moratorium on testing.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

Chapter

I. INTRODUCTION 1

Statement of the Problem 8
Purpose of the Study 9
Hypotheses 11
Significance of the Study 13
Limitations of the Study 13
Definition of Terms 14

II. REVIEW OF THE LITERATURE 18

I. Q. and Reading 18
Cloze Tests and Verbal Intelligence. . 26
Fundamental Factors in Reading Comprehension 27
Wide Range Achievement Test as a Measure of Reading 32
Gilmore Oral Reading Test as a Measure of Reading 37
Durrell Listening-Reading Tests. . . 39
Reliability 40
Sex Differences and Reading Achievement 41

III. TEST ADMINISTRATION, PROCEDURES, AND TECHNIQUES 43

Statistical Analysis 45
Construction 48
Population 49

IV. FINDINGS AND RESULTS 51

Discussion 55
Results 63

Chapter Page

V. SUMMARY, CONCLUSIONS AND IMPLICATIONS . 66

Summary 66
Implications for Future Research . . 69

APPENDIX 71

REFERENCES 98

SAMPLE PAGES OF TESTS

		Page
1.	Wide Range Achievement Test	74
2.	Durrell Listening-Reading Series	76
3.	New York State PEP Test in Reading	82
4.	Gilmore Oral Reading Test	87
5.	Cloze Test	92
6.	Wechsler Intelligence Scale for Children. .	96

CHAPTER I

INTRODUCTION

> Every institution of social significance to society is constantly being evaluated and criticized. Herein lies the seeds of growth and growing excellence. So long as such evaluation is fairly made and the discussion is designed to improve rather than destroy, criticism serves as a constructive end and is to be welcomed.[1]

Every fall since 1965, over one million pupils enrolled in classes in more than five thousand different public and nonpublic schools in New York State are required to take part in a testing program. This Pupil Evaluation Program (PEP) test is given to all pupils in grades 3, 6, and 9 in the areas of reading and mathematics. The rationale for this program is found in the Manual:

> The program was established in September 1965 to help develop and maintain effective allocation, control, and evaluation procedures in the administration and use of ESEA Title 1 funds. It was designed to provide schools and the Department with a single uniform set of statewide test data that would identify educationally disadvantaged pupils

---

[1] B. T. Thayer, Public Education and Its Critics, (New York: MacMillan Co., 1969), p. 5.

> regardless of where they attend school and give an objective picture of the severity of the problems of educational disadvantage in each and every public school system in the State.[1]

The original purpose of providing information "vital to the efficient management of ESEA Title I funds"[2] has been broadened at both the State and local levels. The PEP tests now provide an annual statewide school-by-school inventory of pupil achievement in reading and mathematics. The Manual for the PEP tests states that the results "objectively and reliably identify educational needs and indicate progress in meeting these needs irrespective of whether the funds originate from a local, state, or national source."[3]

The results of these tests are used by State agencies and communities as a basis for statewide educational evaluation, planning and financing. Personnel in school districts also use the test results "to formulate plans and make policy decisions . . . on supervision,

---

[1]The University of the State of New York--The State Education Department Bureau of Pupil Testing and Advisory Services, Albany, New York, New York State Pupil Evaluation Program (Fall, 1973), Directions for Administering, Scoring and Reporting, p. 1.

[2]Ibid.

[3]Ibid.

allocation of personnel, improvement of instruction and assessment of educational quality."[1]

The State has indicated that the only way the program can maintain its value as "a statewide product evaluation and continuous inventory system"[2] is to mandate that every school in the State must administer the Pupil Evaluation Program test to every eligible pupil in its district. The time of administration of this test has also been mandated as falling between the 1st and 12th of October of each school year.

After these tests are administered, they are scored by the classroom teacher at the third grade level and sent to Albany to be scored by machine at the sixth and ninth grade levels. The following April, the sending school receives print-outs of each pupil's performance as well as schoolwide, districtwide, and statewide comparative analyses. The percentage of students performing in the various achievement levels is reported for these various reference groups. Percentile ranks and median raw scores are also reported.

---

[1] Ibid.

[2] Ibid.

The results are reported in terms of nine achievement levels or stanines which may be grouped in three ability ranges. Stanines 1, 2, 3 are the lowest; stanines 4, 5, 6 are equivalent to the middle or average range, and stanines 7, 8, 9 are equivalent to the high range.

These results were given front page coverage in the New York Times and were featured prominently in other papers. Maeroff used this headline to dramatize his article:

> A Solution to Failing Reading Scores Continues to Elude Big City Schools . . . The reading scores of inner city school children sink to new lows each year, and educators and others here and in most big cities have been unable to reverse the trend. A survey of the reading situation by the New York Times has found that the problem--and its seeming intractibility--are common to most big cities, including New York, Chicago, Los Angeles, Milwaukee, Boston and others.[1]

Buskin and Schaeffer in discussing these test score results, said that:

> Standardized test scores in third grade and sixth grade reading and mathematics . . . once again show that those districts with large numbers of disadvantaged students have the worst scores. But this year, a growing number of middle-class districts also have deteriorating scores . . .

---

[1] New York Times, May 27, 1973, p. 1. Gene I. Maeroff, "A Solution to Failing Reading Scores Continues to Elude Big-City Schools."

> While authorities in disadvantaged districts repeated the usual explanations that test scores are affected by large numbers of transient students and a lack of remedial programs, educators in middle-class districts had their own explanations for poor showings on the tests. They ranged from the effect of internal upheaval because of administrative and school board turnover . . . to an increased number of disadvantaged children.[1]

In another article, Buskin felt:

> . . . that only a small handful of districts level with the public and release the results as soon as they are received from the state. But the majority still maintain an unrealistic silence and then hope that parents don't see the scores when they are printed in the daily newspaper.[2]

> . . . Educators contend, with much justification, that too much attention is paid to these scores. Parents use them to compare school districts, to rate teachers, to critique administrators and to abuse school boards.[3]

These test results used to measure reading attainment in the schools yield misleading information to the lay public because they make no mention of the fact that they are designed in such a way that in any "normal" population, fifty per cent of the scores will fall above and fifty per cent below the median which the State established in 1966.

---

[1] Newsday (New York), June 21, 1974, p. 19, Martin Buskin and Tony Schaeffer, "Middle Class Districts Lag on Tests."

[2] Newsday (New York), June 21, 1974, p. 46, Martin Buskin, "Why Treat School Test Scores Like Dirty Secrets."

[3] Ibid.

The median score was established for the PEP test by administering to a wide representative sample and finding the median score. This procedure yields the information that students tested have a comparable relationship to the established median value. If the newly tested students were identical to the normative sample, fifty per cent would be below and fifty per cent above the midpoint on the normative testing.

Kraus, in his book, addressed himself to the potentially harmful effects of using test results in this manner.

> It is as if we were to line up children in the order of their scores, set the norm at midpoint, and then proceed to punish all the children who are below this point--although we used them to establish that midpoint or norm. To many children and parents, scoring below the norm means the child is "sub-normal" . . .
>
> . . . When children are told that they are deficient until they meet the grade norm, the anxiety which is triggered sometimes becomes a paralyzing force and frequently shows its effects until well into the high school years.

In addition to a lack of understanding of what constitutes a norm, the scores reported annually are assumed to be an indication of whether the teachers, the schools,

---

[1]Philip E. Kraus, A Longitudinal Study of Children from Kindergarten into the Adult Years (New York: John Wiley and Sons, 1973), p. 9.

and the students are doing better or worse. The effects have been far reaching and according to Kraus:

> . . . resulted in recent years in odious comparisons between schools, and militant attacks upon teachers and supervisors by parents of children in school with low reading scores.[1]

Shanker, who is co-president of the New York State Teachers Union, was in favor of the publication of these test scores. He felt that the public does have a right to know, but:

> releasing raw scores on tests does not promote understanding. It confuses and stigmatizes. The kind of research and testing that is needed is more difficult and more extensive than what is being done now.[2]

In order to find out what research the State Department had available on the PEP test, a call by the investigator was made to the Bureau of Pupil Testing and Advisory Services, New York State Education Department, in Albany. Mr. Jack A. Maybee,[3] Chief of Pupil Testing and Advisory Services, and Mr. Lott, Chief Bureau of Testing, were asked for any studies that had been done on the PEP test.

---

[1] Ibid, p. 46.

[2] Albert Shanker, "Where We Stand"-New York Teacher, (April 1, 1973), p. 17.

[3] Telephone conversation in November 1973 with Jack A. Maybee, Chief Bureau of Statistics, and Mr. Lott, Chief Bureau of Testing, The University of the State of New York, The State Education Department, Albany, New York.

The investigator was informed that none had been undertaken by the State Education Department. However, for the first time, the Department was requesting those districts that had done any research using the PEP test with other standardized tests to share their findings with them, as soon as they were available.

Over nine million students have been tested since 1965, and the only feedback the Department had has been the scores of these tests.

In 1973, a new reading test for grade three was devised. The district in which the investigator is employed had been subjected to criticism from parents and from the Board of Education because of the PEP test results. It was therefore felt that an evaluation of the new test would prove beneficial to this district.

Statement of the Problem

The Bulletin for the New York State Tests in Reading, Beginning Grades 3 and 6, Form C, states that the Pupil Evaluation Program (PEP) tests are measuring three components of reading comprehension ability:

1) comprehension of explicitly stated material
2) comprehension of implicit or latent meaning, and
3) word knowledge

> Several types of questions are used to measure the pupil's ability to comprehend implicit or latent meaning. Included are questions that measure ability to draw inferences from what is read, derive the meaning of words from context, perceive the main idea or central thought of what is read, and to identify the antecedent of a pronoun.[1]

The issue of interest for this investigation was to find out whether the PEP test was measuring what the State Education Department claimed it to be measuring.

Purpose of the Study

The purpose of this study was to investigate the concurrent validity of the third grade PEP test with standardized tests. The tests chosen for this comparison were The Durrell Reading Test, The Gilmore Oral Reading Test, The WRAT, the Verbal IQ of the WISC, (including the following sub-tests: information, vocabulary, arithmetic, similarities, comprehension and digit span), and the Cloze Test constructed by the investigator.

The investigation of concurrent validity led to several specific questions concerning the relationships among these several measures:

---

[1] Bulletin--Directions for Administering and Scoring New York State Tests in Reading Beginning Grades 3 and 6 Form C, (Albany, New York: Copyright, 1973), p. 5.

A. What is the relationship between the PEP test and reading comprehension tests?

B. What specific intellectual abilities was the PEP test measuring in common with tests of reading comprehension?

C. Was the PEP test measuring the three components of reading comprehension ability the Manual purported it to be measuring?

D. Did the variable of sex have any affect on any measure of the study?

The answers to these questions were explored by measuring the relationship between the PEP test and several variables. These include tests of word recognition and oral and silent paragraph reading. The WISC was utilized to identify those intellectual abilities that correlated highest in a multiple regression. The extent to which the PEP test correlated with oral and silent reading tests and tests of specific intellectual abilities was established, thus demonstrating its concurrent validity with these measures. A correlation matrix, consisting of all of the tests given, was analyzed. The extent to which the PEP test and oral and silent reading tests were measuring the same constructs was demonstrated.

The following null hypotheses were included in the study to test assumptions made regarding the PEP test.

Statistical Hypotheses

1. There is a high positive correlation between the PEP test and the vocabulary sub-test of The Durrell Reading Test.

2. There is a high positive correlation between the PEP test and the comprehension sub-test of The Durrell Reading Test.

3. There is a high positive correlation between the PEP test and the total Durrell Reading Test.

4. There is a high positive correlation between the PEP test and the accuracy sub-test of The Gilmore Oral Reading Test.

5. There is a high positive correlation between the PEP test and the comprehension sub-test of The Gilmore Oral Reading Test.

6. There is a high positive correlation between the PEP test and the reading sub-test of The WRAT.

7. There is a high positive correlation between the PEP test and the score of the Cloze Test.

8. There is a high positive correlation between the PEP test and the scaled score of the information sub-test of The WISC.

9. There is a high positive correlation between the PEP test and the scaled score of the vocabulary sub-test of The WISC.

10. There is a high positive correlation between the PEP test and the scaled score of the arithmetic sub-test of The WISC.

11. There is a high positive correlation between the PEP test and the scaled score of the similarities sub-test of The WISC.

12. There is a high positive correlation between the PEP test and the scaled score of the comprehension sub-test of The WISC.

13. There is a high positive correlation between the PEP test and the scaled score of the digit span sub-test of The WISC.

14. There is a high positive correlation between the PEP test and the total verbal score of The WISC.

15. There is a high positive correlation between the PEP test and the scores obtained by either sex.

Significance of the Study

This study is of potential significance for the lay public, parents, administrators, classroom teachers, students, and the press.

It tested the validity of the PEP test by ascertaining to what extent the PEP test, the dependent variable, is measuring the same factors related to reading and verbal intelligence as are the independent variables.

It may also give insight into the concurrent validity of the PEP test and the factors it is in fact measuring. The determination of these aspects would strengthen the inferences drawn about the PEP test as a measure of reading. These inferences can be drawn by an analysis of all of the components of those tests with which the PEP test has correlation.

Limitations of the Study

1. It was noted that the sample in this study consisted of fifty children from a white, middle-class, suburban school system. Therefore, the generalizations made must necessarily apply to those school districts which have a similar population.

2. Psychology majors in their senior year of college administered The WISC under the supervision of the investigator. The scoring was checked by their Professor. The results are limited to the extent to which the students followed the precise timing, administration, and scoring of The WISC.

3. Many teachers were involved in the administration and hand-scoring of the PEP tests. The results are therefore limited to the extent to which the teachers correctly followed the standardized instructions for the administering and scoring of the PEP tests.

Definition of Terms

1. Statewide Reference Point: The State Reference point for the new Form C Revised third grade reading test is a raw score of fifteen or the twenty-third percentile. Raw scores of fifteen or less represent achievement "below the Statewide Reference Point."[1]

This reference point is important to school districts, inasmuch as ESEA funds are provided for those

---

[1]Pupil Evaluation Program, New York State Education Department, Bureau of Pupil Testing and Advisory Services, February, 1974, letter.

students who fall below this reference point, and are considered "below average."

2. Intelligence Quotient: (I.Q.) scores refer to the results obtained from The WISC. The WISC yields a Verbal IQ, a Performance IQ, and a Full Scale IQ. For this study only the Verbal IQ which yields sub-scores on the following sub-tests: information, comprehension, arithmetic, similarities, vocabulary, and digit span, was used.

3. Concurrent Validity: Concurrent validity is determined by finding the correlation between given scores on the given instrument and scores on some other measures administered at approximately the same time.[1]

4. Regression Equation: A regression equation involves an estimation of one variable (the dependent variable) from one or more related variables (the independent variable).

5. Multiple Stepwise Regression: According to Kerlinger, multiple stepwise regression is:

> The correlation of all the independent variables with the dependent variables are calculated. The independent variable that has the highest zero order

---

[1] C. M. Lindvall, Measuring Pupil Achievement and Aptitudes (New York: Harcourt, Brace & World, 1966), p. 76.

> correlating with the dependent variable is entered first into the analysis. The next variable to enter is the one that produces the greatest increment to R^2, after having taken into account the variable already in the equation. At each succeeding step, the variable with the largest squared semi-partial correlation is the one to enter the equation.[1]

6. Reading Comprehension: According to Spache (who examined a number of analytic studies of reading comprehension and found a high degree of agreement among them), reading comprehension has three factors. The first factor represents the ability to understand word meanings; the second, "probably measures the ability to comprehend the relationships among words and ideas"; and the third factor has been given a variety of labels such as "analysis-synthesis of ideas, or inductive reasoning, or deductive thinking and problem solving, or abstract reasoning." Spache assigned the label "reasoning" to this factor."[2]

7. The Cloze Procedure: The Cloze procedure is:

> primarily a systematic mutilization of a prose passage by deleting, according to some prearranged standard and system, a percentage of words. The

---

[1] Fred M. Kerlinger and Eliza J. Pedhazer, Multiple Regression and Behavioral Research (New York: Holt, Rinehart and Winston, Inc., 1973), pp. 285-286.

[2] George D. Spache, Toward Better Reading (Illinois: Garrard Publishing Company, 1963), pp. 62-63.

mutilated passage is presented to subjects who are instructed to fill in the blanks with the words they feel have been omitted.[1]

8. The Gilmore Oral Reading Tests: The Gilmore Oral Reading Tests are standardized oral paragraphs in reading, graded according to difficulty, which yield scores in Accuracy and in Comprehension.

9. The Wide-Range Achievement Test: The WRAT test yields scores in oral reading, arithmetic and spelling. For the purpose of this study, only the score in oral reading (word recognition) was used.

10. PEP Tests: The PEP test refers to the Pupil Evaluation Program, the standardized test formulated by the State Department in Albany for assessing achievement in Reading and Mathematics in grades 3, 6 and 9.

11. Durrell Listening-Reading Series: The Durrell Listening-Reading Series are standardized silent and listening reading tests that yield scores in vocabulary meanings and comprehension reading of paragraphs arranged according to difficulty. For the purpose of this study, only the silent reading sub-tests in vocabulary, comprehension, and total reading were used.

---

[1]Wilson L. Taylor, "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, XXX (Winter, 1953), pp. 415-33.

CHAPTER II

REVIEW OF THE LITERATURE

The purpose of this chapter is to present a review of related literature and previous investigations. This chapter is divided into three sections. The first section concerns the relationship between IQ tests and reading achievement. The second section concerns the interrelationships among several different measures of reading. The third section concerns the relationship between sex and reading achievement.

IQ and Reading

The relationship of intelligence test scores to reading test scores has been the subject of much research during the past fifty years. According to Neville and Bruininks:

> An examination of past research in this area reveals that reading achievement and intelligence test scores are moderately related (r's from 0.40 to 0.60), that scores on verbal intelligence tests correlate higher with measures of reading than do nonverbal intelligence test scores, and that some

> evidence suggests that the magnitude of correlations between these two attributes increases with advances in chronological age.[1]

It has not been established whether increments with age in correlations could occur due to changes in the test instruments that are used to assess reading and intellectual skills, since the content of group measures in both these areas become more similar at higher age levels. Bruininks, et al, thought that in the initial stages of reading development, "perceptual abilities...may be more important to the development of early word recognition skills than general verbal ability."[2] However, by the second and third grades, reading is less dependent upon decoding skills. The reading task for older children requires more cognitive skills, "(e.g., logical analysis, making generalizations, etc.) as the reading process changes in emphasis from simple word identification to reading comprehension."[3] Towards the end of the primary

---

[1] D. Neville and R. H. Bruininks, Appleton-Century Crofts (New York: H. C. Heywood, Ed.), citing Psychiometric Intelligence, in press.

[2] Bruininks and Bruininks, in press.

[3] Ibid.

grades a child's instructional program reflects this shift away from the mechanics of "learning to read" towards "reading to learn."

Another important factor in evaluating the change in relationships between the IQ and reading test scores with age is the nature of items included in standardized intelligence tests. The Stanford Binet represents a factorial complex measure of intelligence. Heavily weighted in verbal skills, the Stanford Binet in the age range from four to six years nonetheless includes items which purport to test intelligence through tasks which require vocabulary knowledge (pictorial and aural), solution of analogies, verbal comprehension of cause-effect relationships, analysis of likeness and differences, memory for instructions and digits, discrimination of forms, and design copying, etc. Apparently, the tasks included in conventional IQ measures at this age level are related more to skills involved in reading comprehension than to those required in early word identification. Dechant noted:

> The ability to cognize symbolic units, to identify or recognize words, is not predictable by the IQ. The IQ predicts the pupil's ability to take

meaning to the symbolic unit, to cognize semantic units, or to excel in verbal comprehension.[1]

Individual intelligence tests have achieved widespread acceptance as the major instrument in the assessment of academic potential. One of the advantages cited for the Wechsler scales is that they are "minimally influenced by the subject's inability to read."[2] In order to explore whether reading ability was more closely related to survival in high school than IQ, particularly for low achievers, a study was done by Simpson,[3] comparing the effectiveness of individual intelligence tests (WISC and WAIS) to that of individual reading achievement tests (WRAT and Gilmore Oral Reading Tests) in predicting success for low achievers. The sample consisted of 183 graduates and 197 dropouts, a total of 380 students. A reading index was obtained for each student by dividing reading achievement by chronological grade placement.

---

[1] E. V. Dechant, Improving the Teaching of Reading (Englewood Cliffs, New Jersey: Prentice Hall, 1964).

[2] Robert L. Simpson, "Reading Tests versus Intelligence Tests as Predictors of High School Graduation," Psychology in the Schools, (Vol. 7, 1970), pp. 363-365.

[3] Ibid.

Biserial correlations were computed between each of four continuous variables (Verbal IQ, Performance IQ, Full Scale IQ, and Reading Index) and the discrete variable; graduate or dropout. As a result, Simpson concluded:

> The individual reading tests surpass individual intelligence tests in ability to predict successful graduation from regular high school classes for low achieving students. Verbal IQ has moderate value as a predictor of ability to graduate in regular classes, but Performance IQ is practically worthless for this purpose.[1]

Other studies have shown the relationship between the WISC and standardized tests of achievement. Mussen, et al,[2] found moderate correlations between the WISC and the Metropolitan Tests of Reading and Arithmetic and the Stanford Reading and Arithmetic Tests. Stroud, et al,[3] also investigated the relationships between the Full Scale WISC and scores obtained on the Iowa Tests of Basic Skills. Correlations of .66, .66, and .67 were

---

[1] Ibid, p. 364.

[2] Mussen, P., Dean S., and Rosenberg, Margery, "Some Further Evidence on the Validity of the WISC," Journal of Consulting Psychology, (1952, 16), pp. 410-411.

[3] Stroud, J. B., Bloomers, P., and Lauber, Margaret, "Correlational Analysis of WISC and Achievement Tests," Journal of Educational Psychology, (1957, 48), pp. 18-26.

reported with Reading, Arithmetic, and Spelling. They also reported correlations between the three achievement tests and the eleven sub-tests of the WISC.

In a study that compared the prediction of school achievement by a group intelligence test with the WISC, and investigated the long term relationships between WISC and actual school achievement, as well as to study the reliability of WISC scores over a four year period, Conklin and Dockrell found that:

> . . . there is little difference in the predictive qualities of either the Full Scale or the Verbal Scale IQs in relation to achievement. (F.S. Reading Grade 6, .45, Grade 7, .44, Grade 8, .67, Grade 9, .52; V.S. Reading Grade 6, .48, Grade 7, .43, Grade 8, .62, Grade 9, .56; P.S. Reading Grade 6, .28, Grade 7, .32, Grade 8, .54, Grade 9, .36.)[1]

The Full Scale and the Verbal seem to predict better than the Performance Scale IQs. However, these differences, although approaching significance, are not statistically significant ($p > .05$).

The reliabilities over a four year period follow:

---

[1] R. C. Conklin and W. B. Dockrell, "The Predictive Validity and Stability of WISC Scores over a Four Year Period," Psychology in the Schools, (Vol. 4, July 1967), pp. 263-266.

Sub-test	Inf.	Com.	Ari.	Sim.	Voc.	P.C.	P.A.	B.D.	O.A.	Cod.
1961-1965	41	27	51	34	58	41	13	53	40	59

The Vocabulary sub-test at grade 9 level appeared to be a consistently good predictor, and the Arithmetic sub-test, the best predictor at grade 6. The investigators concluded that at the "lower grade level a numerical abstract ability is involved in academic achievement, whereas at the higher grade level a verbal abstract ability is involved in all types of academic achievement."[1] The investigators were not sure whether this change was a "function of the curriculum" or reflects what Piaget (Inhelder and Piaget),[2] referred to as "a progress from concrete to formal thinking."

In a study investigating the relationship between WISC results and classroom comprehension as stated by teachers, Naiven and Bierbryer used thirty male and seven

---

[1] Ibid.

[2] Inhelder, B. and Piaget, J., The Growth of Logical Thinking from Childhood to Adolescence, (London: Routledge and Kegan, Paul, 1958).

female first through sixth grade children attending suburban Long Island and New Jersey public elementary schools. The subjects had been referred to the school's psychologist for many reasons including underachievement and behavioral difficulties. Each child was administered the WISC by the school psychologist and was rated by the classroom teacher on the Comprehension measure of the Devereux Elementary School Behavior Rating Scale. They concluded:

> . . . The WISC is a valid predictor of elementary school children's classroom comprehension. Despite its name, the Comprehension Sub-test is apparently not the best predictor among the scores provided by the WISC. Rather, the Information and Vocabulary Sub-tests and Verbal IQ seem to be the most effective in this regard.[1]

Robinson did a study of thirty seriously retarded readers referred to the University of Chicago Reading Clinic for diagnosis in order to study the causal factors responsible for their reading disability. There was a psychometric examination, achievement and reading scores were obtained, preference tests given, social

---

[1]Frederick B. Naiven and Bruce Bierbryer, "Predicting Elementary School Children's Classroom Comprehension from Their WISC Results," Journal of Clinical Psychology, Vol. XXV, No. 1 (January, 1969), pp. 75-76.

history recorded, psychiatric study, pediatric examination, neurological, visual, auditory and speech examinations and endocrine examination. Group conferences were held on each subject.

The criteria was an IQ of at least 85 on the New Stanford Binet. The chronological ages ranged from 6 years to 15 years. The reading retardation ranged from 0.9 grades to 6.4 grades and median retardation was 3.9 grades.

Robinson concluded: "There is a positive relationship between intelligence scores and reading test scores."[1]

Cloze Tests and Verbal Intelligence

The Cloze Test is a reading test which measures the subject's ability to deal with relationships among words. The relationship of Cloze scores to measures of verbal intelligence and general intelligence has also been demonstrated. Results obtained by Cloze Test utilizing the random any word (or Nth word) deletion systems indicate that there is a definite relationship

---

[1] Helen M. Robinson, Why Pupils Fail in Reading (Chicago: University of Chicago Press, 1966).

to tests of verbal intelligence and general intelligence. Taylor,[1] et al, found a relationship of significant magnitude of .6 to .7. The only study that did not find a high correlation between Cloze Tests and reading achievement was the study of Weaver and Kingston[2] who found a loading of .73 on Verban Comprehension, .32 on Cloze and .34 on the rote memory factor. Horton, in analyzing the validity of the Cloze Test with other measures, found:

> It would appear that the variances shared among Cloze Tests, reading comprehension tests, reading gain tests, and verbal intelligence tests are probably a measure of the subject's ability to deal with the relationships among words and ideas. Thus, it is probably this common variance which accounts for the concurrent validity of Cloze procedure with these other measures.[3]

Fundamental Factors in Reading Comprehension

The PEP test purports to be measuring three components of reading: (1) comprehension of explicitly

---

[1] Wilson L. Taylor, "Cloze Readability Scores as Indices of Individual Differences in Comprehension and Aptitude," Journal of Applied Psychology, XXXXI (February, 1957), pp. 25-26.

[2] Wendell W. Weaver and Albert J. Kingston, "A Factor Analysis to the Cloze Procedure and Other Measures of Reading and Language Ability," The Journal of Communication, XIII (December, 1963), pp. 254-61.

[3] Raymond J. Horton, "The Construct Validity of Cloze Procedure," (unpublished, Ph.D. Dissertation, Hofstra University, 1973), p. 162.

stated material, (2) comprehension of implicit or latent meaning, and (3) word knowledge.[1]

Lennon, in an article entitled "What Can Be Measured," discussed studies in this area and concluded that there are only six factors which account for more than ninety percent of reading competence. They include the three components of reading comprehension the State Education Department is measuring, plus a fourth, called appreciation. Lennon stated:

> It seems entirely clear that numerous superficially discrete reading skills to which separate names or titles have been attached are, in fact, so closely related, as far as any test results reveal, that we must consider them virtually identical.
>
> It would seem that we may recognize and hope to measure reliably the following components of reading ability: 1) a general verbal factor, 2) comprehension of explicitly stated material, 3) comprehension of implicit or latent meaning, and 4) an element that might be termed appreciation.
>
> The "verbal factor" in this context is intended to connote word knowledge; breadth, depth, and scope of vocabulary... Every investigation shows vocabulary to be substantially related to other measures of reading ability...[2]

---

[1]Bulletin--Directions for Administering and Scoring New York State Tests in Reading Beginning Grades 3 and 6 Form C (Albany, New York: Copyright, 1973), p. 5.

[2]Robert T. Lennon, "What Can Be Measured," Reading Teacher, Vol. 15, No. 5 (March, 1962), pp. 326-327.

Under comprehension of explicitly stated material, Lennon included such skills as locating information, following directions, and comprehending literal meanings. Under comprehension of implicit meaning, he included making inferences, predicting outcomes, using context clues, getting the main idea and all of those outcomes that are labeled reasoning in reading.

Traxler,[1] in 1941, sought to ascertain whether the several parts of the Van Wagenen-Dvorak Diagnostic Examination of Silent Reading Abilities yielded measures which were independent enought to warrant their separate measurement and use as a basis for diagnostic and remedial work. After studying the results of 116 tenth grade students, he concluded that:

> . . . measures of central thought, clearly stated details, interpretation, integration of dispersed ideas and ability to draw inferences, appear to be measuring closely related reading abilities.[2]

Traxler found most of the parts so highly correlated that diagnosis based on the scores had little

---

[1]Arthur E. Traxler, "A Study of the Van Wagenen Dvorak Diagnostic Examination of Silent Reading Abilities," Education Records Bureau Bulletin, No. 31 (January, 1941), p. 33.

[2]Ibid, p. 33.

real meaning. In fact, when the intercorrelations were corrected for attenuation, most approached unity.

Davis sought to identify the fundamental factors in reading comprehension and to provide a means of measuring them. He developed a test of nine supposed categories and computed the intercorrelations among the nine tests. Davis interpreted a factor analysis of the results as indicating the presence of nine factors, six of them clearly significant. These latter included:

> . . . word knowledge; ability to manipulate ideas and concepts in relation to one another--"reasoning in reading"; ability to grasp the author's expressed ideas; ability to identify the writer's intent of purpose; ability to follow the organization of a passage, and knowledge of literary devices and techniques.[1]

Of Davis' nine factors, word knowledge accounted for by far the greatest part of the variance followed by the so-called "reasoning in reading" and the literal meaning factors.

Hall and Robinson reported an attempt to develop independent measures of various aspects of reading. They developed a battery that included twenty-five measures.

---

[1]Frederick B. Davis, "Fundamental Factors of Comprehension in Reading," (unpublished, Ed.D. dissertation, Howard University, 1941), and in Psychometrika 9 (1944), pp. 185-187.

Factor analysis of the results of administration of this battery of tests to one hundred college students yielded six factors, which Hall and Robinson defined as:

> "attitude of comprehension accuracy," an "inductive" factor, a "verbal or word-meaning" factor; a "rate for unrelated facts" factor; a "chart reading" factor; and a sixth "undefined" factor.[1]

Six factors were sufficient to account for the variance in the twenty-five separate reading measures.

Hunt, in 1952, did a study to determine whether the six factors identified by Davis would reappear in an independent investigation, or more generally whether reading comprehension was made up of aspects sufficiently specific to be measurable as independent variables. He found:

> In general the items classified in any given area--as for example vocabulary items, or items intended to measure reasoning, correlated no higher with the total score on the ability they were supposed to measure than with the score on any of the other abilities. That is to say, there was no evidence in the item-discrimination statistics that the items possessed any differential validity as

---

[1] W. E. Hall and F. P. Robinson, "An Analytical Approach to the Study of Reading Skills," Journal of Educational Psychology, XXXVI (1945), pp. 429-442.

> measures of one aspect of reading comprehension rather than another. Factor analysis of the scores led to the same general conclusion--that, except for the vocabulary test, the other measures "may be measuring much the same function of reading comprehension." Reasoning and word knowledge merge as the first and second factor in this research.[1]

Wide Range Achievement Test -- A Measure of Reading

The WRAT was first published in 1946. Its use by school psychologists, teachers, and reading clinics has steadily increased, due to its ease of administration, and range of ability (kindergarten to superior adult). In spite of the fact that it is "the second most frequently used achievement test in psychological clinics;"[2] a review of the literature revealed only three studies (one of which was unpublished) that have been done pertaining to the validity of the WRAT.

Garlock, et al, did a study of 108 atypical subjects in grades one through twelve of which thirty-five were in EMR (educable mentally retarded classes) in order

---

[1] J. M. Hunt, cited in "What Can Be Measured," Reading Teacher Vol. 15, No. 5 (March, 1962).

[2] Norman D. Dundberg, "The Practice of Psychological Testing in Clinical Services in the United States," American Psychologist, Vol. 16 (1960), pp. 79-83.

to determine the validity of the WRAT with this population. The WRAT and The Gilmore Oral Reading Tests (Form A) were administered by seven school psychologists. The investigators determined the correlations between tests by "holding grade constant (r12. grade) in order that the validity coefficient to be expected within a single grade could be estimated."[1] They found that even when grade level is held constant, eighty percent of the variance from the WRAT is overlapping with the Gilmore Accuracy variance. When the correlation was corrected for attenuation, the investigators found "a value of .96 shows the WRAT to convey essentially all the reliable information found on the Gilmore Accuracy, hence supporting its validity."[2]

The correlation of the WRAT with other reading criteria was found to be high at grade 5, the correlation with the Sangren-Woody Silent Reading Test was .78; with the New Stanford Reading Test .80; with Teacher Ratings .79; with Semester Reading Grades .88; with CAT-Reading

---

[1] Jerry Garlock, Robert S. Dollerhide, and Kenneth D. Hopkins, "Comparability of Scores on the Wide Range and the Gilmore Oral Reading Tests," California Journal of Educational Research, Vol. 16 (March, 1965), pp. 54-57.

[2] Ibid, p. 55.

Vocabulary .67; Comprehension .67, and total Reading .71. The correlation of adult mentally retarded on the Gray's Oral Reading Paragraphs was .94.

The Comprehension test of the Gilmore was shown to be only moderately related to the other two, yet it contained a substantial amount of information not supplied by the other two. Garlock,[1] et al, found the WRAT and the Accuracy sub-test of the Gilmore Oral Reading test conveyed almost identical and interchangeable information. The standard errors of estimate was only .66 grade equivalents. Garlock felt that:

> . . . the relationships of both of the above tests with the comprehension section of the Gilmore was substantial yet low enough to clearly illustrate the need for both types of tests for a complete individual appraisal. The data supported the hypothesis that the WRAT is a valid measure of reading accuracy and/or vocabulary even with the atypical populations of referred and mentally retarded pupils.[2]

Another study to evaluate the reading test of the WRAT in terms of its concurrent and congruent validity was done by Hopkins, et al. The concurrent validity of the WRAT was assessed by correlating WRAT results with

---

[1] Ibid, p. 56.

[2] Ibid, p. 56.

independent ranking of reading ability by the classroom teacher. The sample consisted of 502 students in grades one through five in schools where means have closely paralleled the national means in achievement and intelligence on previous standardized test results. Hopkins concluded that:

> Although the WRAT correlated highly (.7 to .8) with teacher rankings at all grade levels investigated, the coefficients in grades 3, 4, and 5 were higher than those in grades 1 and 2. It is the writer's opinion that this results, in part, from increased criterion reliability resulting from increased reading variance by grade.
>
> It is interesting to note that the WRAT, which has no comprehension section, did correlate as highly with the RC (Reading Comprehension) Section of the CRT (California Reading Test) as it did with the RV (Reading Vocabulary) section...Both the WRAT and the CRT possessed substantial concurrent validity.[1]

Oscar Krisen Buros has edited a number of yearbooks which contain reviews of test instruments by outstanding authorities in various fields. In a review of the WRAT, Merwin called attention to the fact that the authors of the test "report questionably high reliability

---

[1]Kenneth D. Hopkins, James C. Dobson and O. A. Oldbridge, "The Concurrent and Congruent Validities of the Wide Range Achievement Test," Educational and Psychological Measurement, (Vol. XXII, No. 4), 1962, pp. 791-793.

coefficients by one age group."[1] He felt that:

> . . . it wasn't possible to determine the extent to which the reported split-half coefficients based on "odd-even" scores are affected by the sequential administration and scoring procedures used for the test, from the information provided.[2]

Merwin also questioned the procedures used and the reported magnitude (e.g., low of .981 out of 14 reliability coefficients on the reading test). He found that it was very difficult to assess the value of the validity information the authors provided in the manual, inasmuch as it was unclear which of the 1965 two-level editions, the authors were basing their results on. The one positive thing the investigator found was that it was "possibly a potentially useful clinical tool for the psychologist working with specialized cases, but for general school use it is impractical."[3]

Thorndike, in his review, found the WRAT of "value in a clinical or a research setting in which one is testing individually persons of diverse abilities and

---

[1] Jack Merwin, "The Wide Range Achievement Test," The Seventh Mental Measurement Yearbook, Oscar Krisen Buros, editor (Highland Park, N. J., 1972) Vol. 1, p. 36.

[2] Ibid, p. 36.

[3] Ibid, p. 37.

back ground"[1] and needed to get a quick estimate of their general level of ability and educational background. He would not recommend it for other purposes.

Gilmore Oral Reading Test As a Measure of Reading

Harris, in reviewing this test, stated:

> Evidence of care in constructing the paragraphs is shown in data that indicate fairly regular increases in the paragraphs in length, word difficulty, number of polysyllabic words, mean number of words per sentence, and percentage of complex sentences. The content seems somewhat more appropriate for non-disadvantaged children than for disadvantaged children. . . . The accuracy score is the most important of the three Gilmore scores. It is a measure of fluency as well as of accuracy of word recognition since hesitations, repetitions, disregard of punctuation and self-corrected are all counted as errors. . . .
>
> The Gilmore Comprehension score is based on the answers to five questions per paragraph. This seems to be mainly a measure of short-time recall of directly stated details, with very few questions calling for judgments or inferences. It therefore differs considerably from the kind of comprehension measured by most silent reading tests. It also has relatively poor reliability. Thus the Comprehension score is disappointing, in terms of both content validity and reliability. . . .
>
> Despite its shortcomings, the updated Gilmore is among the best standardized test of accuracy in oral reading of meaningful material now available.

---

[1] Robert L. Thorndike, "The Wide Range Achievement Test," The Seventh Mental Measurement Yearbook, Oscar Krisen Buros, editor (Highland Park, N. J.: The Gryphon Press, 1972), Vol. 1, p. 36.

The usefulness of the Comprehension and Rate scores is more questionable.[1]

Smith questioned the ability of the Gilmore to measure comprehension ability with an oral test:

> . . . One does not normally read orally, for his own comprehension, but rather for the benefit of listeners. The act of oral reading requires only the most superficial comprehension; therefore the skilled oral reader has his concentration elsewhere. If he wished to read for his own comprehension, he usually would read silently. The appropriateness of testing a silent reading act (comprehension) with a test of oral skill (fluency, etc.) is extremely doubtful. . . .
>
> The test of accuracy (actually oral word-calling) should be useful in word recognition diagnosis if types of errors are analyzed. However, the scoring practice of weighting different types of "errors" equally seems questionable. . . .
>
> . . . the test requires individual administration and is rather time consuming, and the results have limited use in view of the invalid assumptions upon which the test is based. It would therefore seem that the time might be much better spent on a well-constructed informal reading inventory which tests oral reading with an oral test and reading comprehension with a silent test, and which samples types of comprehension, not simply recall of details.[2]

---

[1]J. Albert Harris, "The Gilmore Oral Reading Test," The Seventh Mental Measurement Yearbook, Oscar Krisen Buros editor, (Highland Park, N. J.: The Gryphon Press, 1972), p. 736.

[2]Kenneth B. Smith, "The Gilmore Oral Reading Test," The Seventh Mental Measurement Yearbook, Oscar Krisen Buros editor, (Highland Park, N. J.: The Gryphon Press, 1972), pp. 737-738.

Durrell Listening Reading Series

Bormuth questioned some of the test items in this test but nevertheless felt that:

> The content validity of the Durrell Listening Series is outstanding among standardized vocabulary and comprehension tests . . . these reading and listening tests represent unusually sophisticated work in the selection of a model for emulation by other test makers. However, the tests lose some of their value because of the designs of the test items . . . But the reviewer believes that each of the reading and listening tests is useful in its own right and compares very favorably with other available tests of reading and listening abilities.[1]

Spache, in his review of the same test, stated:

> . . . construct validity, as measured by the readability levels of the selections or by other measure of the difficulty of the content, and concurrent validity, as determined by correlations with comparable tests, remain weak.
>
> Perhaps the best use of this latest version of parallel listening-reading tests would be as a research instrument to explore those questions for which the authors assume they have the answers . . . As it is now, we have no real evidence to justify the basic assumptions made by the authors in preparing the series.[2]

---

[1]John R. Bormuth, "Durrell Listening Reading Series," The Seventh Mental Measurement Yearbook, Oscar Krisen Buros, (Highland Falls, N. J.: The Gryphon Press, 1972), pp. 727-728.

[2]George D. Spache, "Durrell Listening Reading Series," The Seventh Mental Measurement Yearbook, Oscar Krisen Buros, editor (Highland Falls, N. J.: The Gryphon Press, 1972), pp. 728-729.

A review of related research did not reveal any studies that had been done on the Pupil Evaluation Program (PEP) test. A telephone call to the State Department of Education corroborated this fact.[1]

Reliability

According to Jastak and Jastak:

> . . . Many methods have been devised to study the reliability of achievement. Correlating parallel forms of a test, repeat administrations of the same test, split-half forms of one test and calculating coefficients by formulas based on the analysis of item difficulty or variance are among the most common procedures.[2] Guilford[3] cautions that reliability should be "the minimum information provided concerning a test and that it is not to be considered the most important or most useful information."

In determining the size of reliability coefficien , Jastak and Jastak felt the overriding factors are primari y statistical in nature, regardless of the method used.

---

[1]Telephone conversation with Jack A. Maybee in November, 1973, Chief, The University of the State of New York, The State Education Department, Albany, New Yor .

[2]Examiner's Manual, The Wide Range Achievement Te t, J. F. Jastak and S. R. Jastak Guidance Association (1965)

[3]J. P. Guilford, Psychometric Method (New York: McGraw Hill Book Co., 1934), p. 75 cited in Examiner's Manual, The Wide Range Achievement Test, J. F. Jastak and S. R. Jastak Guidance Associates (1965).

> Statistical reliability . . . is a function of the following conditions: the number of items in a test, the time it takes to complete the test, the population sampling used, the range of ability tapped, the specific technique of comparing or splitting two forms, the manner of scaling of the item within the test, the method of answering the test items, the unit of measurement, and the dispersion of item difficulties within a scale.[1]

Sex Differences and Reading Achievement

Investigations have been conducted over the years to determine whether sex is an influential factor in the reading process.

Miller investigated the scores of 541 boys and 491 girls (of the same chronological age) on the Stanford Achievement Test. They were in grades two through seven. Miller reported:

> Comparisons made between the sexes when all the boys and girls of one yearly age are considered as a unit, show the median scores of the boys to be higher more often than those of the girls. Comparisons of the means of the same age group show that the girls rank higher than the boys. This study shows no marked superiority in achievement on the part of either sex.[2]

---

[1] Ibid.

[2] William A. Miller, "Achievement Scores of Boys and Girls of the Same Age," Elementary School Journal, XXXII (May 1932), p. 676.

Wilson did a study over a three year period of reading progress in kindergarten through third grade. He tested 400 children of average or above mental ability as measured by the Stanford-Binet Intelligence Scale, using three types of the Gates Reading Test, the Stanford Achievement Test and the Progressive Achievement test. He reported that:

> It would seem that superiority in reading ability tended to be accompanied by superiority in the important letter abilities functioning in the mechanics of reading rather than with mental superiority or with sex in and of themselves.[1]

The findings of Broom with regard to sex and reading agreed with Wilson. Broom used the El Paso Silent Reading Tests, Form 1 and 11, to compare the reading comprehension ability among 9,276 Anglo-American and Spanish American pupils in grades five through eight. His conclusion was that: "Sex differences were found to be not statistically significant."[2]

---

[1] Fran I. Wilson, "Sex Differences in Beginning Reading in a Progressive School," Journal of Educational Research, XXXII (April, 1959), p. 570.

[2] M. E. Broom, "A Study of the Race and Sex Difference in Reading Comprehension," Journal of Educational Research, XXXIV (April, 1941), p. 593.

CHAPTER III

TEST ADMINISTRATION, PROCEDURES AND TECHNIQUES

Chapter III is concerned with the test administration, procedures and techniques, population, the construction of the tests, and statistical analysis.

In September, sixty third-grade students enrolled in the Old Bethpage Elementary School were selected for the study.

Using a table of random numbers, this sample of sixty subjects was randomly and independently drawn from the entire third grade population of 130 students. Due to absences, ten subjects did not complete the entire battery of tests and/or make-up tests and were dropped, leaving a sample of fifty subjects.

The Wechsler Intelligence Scale for Children (WISC) was administered during the period of October, November, December, and February of 1973-74. The tests were administered by psychology majors of Queens College. They were supervised by the investigator and the scores were checked by their professor.

The Vocabulary and Comprehension Reading Subtests of the Durrell Listening-Reading Series, Intermediate Level, Form DE, were administered in two sessions by the investigator according to the directions in the manual during the month of February, 1974. This form was selected by the investigator. The Primary Battery ranges from Grades 1.0 to 3.5, and the Intermediate Battery from grades 3.5 to 6.0. There are three overlapping levels of the Durrell tests, each level having a distinct set of items. The authors established a continuous scale through an overlapping program. In this program, each battery was administered to the grade levels for which it was intended as well as to one grade above or below the appropriate grade levels. Grade equivalents for values above and below empirically determined points were obtained by extrapolation.

The Cloze Test was administered by the investigator to each third grade class. The tests of those subjects involved in the study were then selected and scored.

The Wide Range Achievement Test was administered individually by the investigator during the months of October through February.

The Gilmore Oral Reading Test was administered by the investigator individually during the months of October through February.

The Pupil Evaluation Program (PEP) test was administered by each third-grade classroom teacher to the entire class. The tests of those subjects involved in the study were scored by the teacher and incorporated in this study.

All of the tests administered by the investigator were scored by the investigator according to the directions in the manual. The raw score total was the score used in all of the tests, with the exception of the WISC. For the WISC, the Verbal sub-test Scaled Scores, and the Verbal IQ were used.

Statistical Analysis

Three multiple stepwise regression analyses were computed using the Statistical Package for Social Sciences-Version of March 13, 1971. The varaibles used in the three regression analyses were:

1 Dependent Variable PEP

2 Vocabulary Durrell

3 Comprehension Durrell

4 Total Durrell

5 Accuracy Gilmore

6 Comprehension Gilmore

7 Wide Range Achievement Test

8 Cloze

#10 Information Scaled Score WISC

#12 Comprehension Scaled Score WISC

#14 Arithmetic Scaled Score WISC

#16 Similarities Scaled Score WISC

#18 Vocabulary Scaled Score WISC

#20 Digit Span Scaled Score WISC

#22 Total Scaled Score WISC

#24 IQ

#25 Sex

The variables used in the first multiple regression analysis were:

1 Dependent Variable PEP

2 Vocabulary Durrell

3 Comprehension Durrell

5 Accuracy Gilmore

6 Comprehension Gilmore

7 Wide Range Achievement Test

#10 Information Scaled Score WISC

#12 Comprehension Scaled Score WISC

#14 Arithmetic Scaled Score WISC

#16 Similarities Scaled Score WISC

#18 Vocabulary Scaled Score WISC

#20 Digit Span Scaled Score WISC

These are the variables used in the second multiple regression analysis:

1 Dependent Variable PEP

4 Total Durrell

5 Accuracy Gilmore

6 Comprehension Gilmore

7 WRAT

8 Cloze

#24 IQ

These are the variables used in the third multiple regression analysis:

1 Dependent Variable PEP

2 Vocabulary Durrell

3 Comprehension Durrell

5 Accuracy Gilmore

6 Comprehension Gilmore

7 WRAT

8 Cloze

#10 Information Scaled Score WISC

#12 Comprehension Scaled Score WISC

#14 Arithmetic Scaled Score WISC

#16 Similarities Scaled Score WISC

#18 Vocabulary Scaled Score WISC

#20 Digit Span Scaled Score WISC

#25 Sex

Means, standard deviations and correlation coefficients were computed for all tests administered.

An intercorrelation matrix, using pearson-product moment correlation for all tests used in the analysis, was computed.

The multiple R, R^2, standard error of R, analysis of variance, regression, residual, and mean square, were computed for each test. The multiple correlations gave the respective degree of relationship that existed between all of the variables in the three respective equations.

The reliability of the PEP test was computed by the investigator by the split-half method. The odd-item total scores were correlated with the even-item total

scores. The correlation was corrected by the Spearman-Brown correction formula. The reliability was .91.

The program for the correlation matrix of the half scores for statistical measures used by the investigator to compute the reliabilities was the Pearson-Product Moment Basic Program Stat. 12 Source, Dartmouth College Computes Corr. N series of data.

Population

Central School District #4, of which the Old Bethpage School is a part, is composed of two communities--Plainview and Old Bethpage. It is an unincorporated area in the Town of Oyster Bay in eastern Nassau County on Long Island. The school district's existing public school plant facilities are two high schools, two junior high schools, four intermediate schools, and four primary schools. In addition, three parochial schools are maintained. Within the next two years, four of the existing primary schools will be closed due to declining school population.

The community was described by Eisner in 1968.[1] It consists of about 9,000 one-family residences.

---

[1] Joseph Eisner, "The Plainview-Old Bethpage Community--A Survey," (unpublished, Plainview-Old Bethpage Library, Plainview, 1968, mimeographed).

Census information, which can be seen in Table 6, lists the total population in 1970 as 33,188.

The school tax of $17.69 per $100.00 assessed valuation was the highest in Nassau County and the State of New York in 1972.

According to the United States Census reports,

> The median income for Plainview is $16,198 and for Old Bethpage it is $17,933. Of the seventy-eight communities in Nassau County, Plainview ranks twenty-six and Old Bethpage ranks nineteen in median income. Of the 145 communities in Nassau and Suffolk Counties, Plainview ranks thirty-three and Old Bethpage ranks twenty-two in median income.[1] (See Table 6)

---

[1] Plainview-Old Bethpage Federation of Teachers, Negotiating Bulletin Update (1972), p. 2 (mimeographed).

CHAPTER IV

FINDINGS AND RESULTS

Chapter IV is divided in three sections. Section one reports the findings of this investigation. Section two presents the discussion and Section three the results of this investigation.

Section One

Table 1 lists the means and standard deviations for all of the variables included in the analysis.

Table 2 represents a matrix of intercorrelations which lists all of the variables that went into the analysis and their intercorrelations with the PEP test and with each other.

Table 3 lists those variables that accounted for additive increments in the variance of the dependent variable, the PEP test. It indicates that the variables that accounted for the highest correlation with the PEP test were the WRAT, the Gilmore Comprehension Test and the Similarities sub-test of the WISC.

TABLE I

Means and Standard Deviations of Tests Administered
Dependent Variable PEP Test

N50

Name of Test	Means	Standard Deviations
PEP	32.6200	10.4156
Vocabulary Durrell	36.8800	10.9705
Comprehension Durrell	24.2400	8.7191
Total Durrell	59.0000	17.4917
Accuracy Gilmore	37.7400	11.1370
Comprehension Gilmore	27.2000	8.9191
WRAT	65.5200	8.5243
Cloze	22.7800	6.4659
WISC Information R.S.*	11.4000	2.0894
WISC Information S.S.**	11.6000	2.8428
WISC Comprehension R.S.	10.7200	3.1170
WISC Comprehension S.S.	11.4200	3.5114
WISC Arithmetic R.S.	8.2800	1.5256
WISC Arithmetic S.S.	12.2200	2.7796
WISC Similarities R.S.	11.3200	3.6670
WISC Similarities S.S.	14.1200	3.2112
WISC Vocabulary R.S.	29.0400	7.3816
WISC Vocabulary S.S.	11.8000	3.1168
WISC Digit Span R.S.	8.9200	1.7710
WISC Digit Span S.S.	10.6800	3.2853
WISC Total Verbal R.S.	71.9000	22.5427
WISC Total Verbal S.S.	70.8200	13.1625
WISC Verbal IQ R.S.	55.8400	18.2962
WISC Verbal IQ S.S.	111.9400	13.8997
Sex	0.5400	0.5035

*R.S. Raw score
**S.S. Scaled score

TABLE II

CORRELATION MATRIX OF TESTS
ADMINISTERED IN THE INVESTIGATION

Var. #	Name of Test	PEP	DURRELL VOC	DURRELL COMP	DURRELL TOTAL	GILMORE ACC	GILMORE COMP	WRAT	CLOZE	WISC INFO	WISC COMP	WISC ARITH	WISC SIM	WISC VOC	DIGIT SPAN	VERBAL IQ	SEX
01	PEP	100	595	559	472	600	552	717	510	443	296	325	350	575	197	547	100
02	Voc.)		100	618	644	571	407	697	476	389	266	319	210	434	240	437	015
03	Comp)Durrell			100	568	501	385	527	651	351	296	322	232	462	219	518	002
04	Total)				100	403	209	554	467	276	126	196	070	340	235	331	144
05	Acc.)					100	249	819	435	205	032	235	016	463	178	285	055
06	Comp)Gilmore						100	384	480	415	249	365	300	479	371	535	-056
07	WRAT							100	508	376	214	318	070	549	314	454	-029
08	Cloze								100	326	302	294	153	427	218	470	-157
010	Information									100	410	688	466	451	148	765	-054
012	Comprehension										100	273	385	496	244	728	-039
014	Arithmetic											100	196	375	325	665	147
016	Similarities												100	408	386	598	-003
018	Vocabulary													100	103	721	096
020	Digit Span														100	453	230
024	Sex															100	136
025	IQ																100

TABLE III

MULTIPLE STEPWISE REGRESSION
CRITERION VARIABLE - PEP TEST
PREDICTOR VARIABLES - WRAT, CLOZE, SUB-TESTS OF
THE WISC VERBAL, GILMORE AND DURRELL

Var. #	Name of Test	Multiple R	R^2	R^2 Change	Simple R
07	Wide Range Ach. Test	717	514	514	717 *
06	Comprehension Gilmore	777	604	090	553
16	Similarities WISC S.S.	808	653	049	350
03	Comprehension Durrell	816	665	012	559
20	Digit Span	823	677	011	197
05	Accuracy Gilmore	823	677	000	596
12	Comprehension WISC S.S.	824	678	000	296
18	Vocabulary WISC S.S.	824	678	000	575
02	Vocabulary Durrell	824	678	000	596

* Decimal points are omitted.

The following tests were not included in the analysis since they did not significantly increase the variance accounted for in the dependent variable.

08	Cloze
10	Information Scaled Score
14	Arithmetic Scaled Score

Table 4 lists those variables that accounted for additive increments in the variance of the dependent variable, the PEP test, when Verbal IQ was added. Its contribution to the variance was .016 percent, which was not significant.

Table 5 lists those variables that accounted for additive increments in the variance of the dependent variable, the PEP test, when sex was added. Its contribution to the variance was -.100 percent which has no significance.

Section Two

Discussion

From an analysis of the matrix of intercorrelations, it was concluded that the test that related most significantly with the PEP test was the WRAT, a test of oral word recognition. The simple r was .716 and the r^2 was .513. Garlock found this test "a valid measure of reading accuracy and/or vocabulary."[1] Hopkins[2] found that this test (which has no comprehension section) correlated as highly with the Reading Comprehension section

---

[1]Garlock, California Journal of Educational Research, Vol. 16, 1965, pp. 54-57.

[2]Hopkins, Educational and Psychological Measurement, XXII, 1962, pp. 791-793.

TABLE IV

MULTIPLE STEPWISE REGRESSION
CRITERION VARIABLE - PEP TEST
PREDICTOR VARIABLES - WRAT, CLOZE, WISC VERBAL IQ,
AND SUB-TESTS-GILMORE

Var. #	Name of Test	Multiple R	R^2	R^2 Change	Simple R
07	WRAT	717	514	513	717 *
06	Comprehension Gilmore	777	604	090	553
24	IQ	788	621	017	547
04	Total Durrell	792	627	006	472
05	Accuracy Gilmore	796	633	006	596
08	Cloze	796	633	000	510

* Decimal points are omitted.

TABLE V

MULTIPLE STEPWISE REGRESSION
CRITERION VARIABLE - PEP TEST
PREDICTOR VARIABLES - WRAT, CLOZE, SUB-TESTS OF
THE WISC VERBAL, GILMORE, DURRELL, AND SEX

Var. #	Name of Test	Multiple R	R^2	R^2 Change	Simple R *
07	WRAT	717	514	514	717
06	Comprehension Gilmore	774	604	090	552
16	Similarities WISC	808	653	049	350
03	Comprehension Durrell	816	665	012	559
20	Digit Span	823	677	011	197
25	Sex	824	678	001	-100
05	Accuracy Gilmore	824	679	000	596
12	Comprehension WISC	824	680	000	296
08	Cloze	825	680	000	509

* Decimal points are omitted.

The following tests were not included in the analysis since they did not significantly increase the variance accounted for in the dependent variable.

02	Vocabulary, total Durrell
10	Information Scaled
14	Arithmetic
18	Vocabulary

of the California (silent) Reading Test as it did with the Reading Vocabulary Section. In Lennon's[1] investigation on "What Can Be Measured," he found that "vocabulary was substantially related to other measures of reading." The latter references suggest that the Word Recognition portion of the WRAT may have reference to a more global set of reading performance than simple word calling.

The test that correlated next highest to the PEP test was the Comprehension test of the Gilmore Oral Reading Test, a test of oral comprehension. The simple r was .553 and the r^2 increased to .604 when the Gilmore was added to the WRAT. The increase of variance was .091 percent. Harris,[2] felt the questions in this test seemed to be mainly a "measure of short-time recall of directly stated details. . . ." The PEP Test (a test of silent comprehension) employs this same short-time recall technique throughout.

In a study comparing the Wide Range Achievement Test and the Gilmore Oral Reading Test, Garlock, Dollarhide and

---

[1] Lennon, Reading Teacher, 1962, Vol. 15, pp. 326-327.

[2] Harris, The Seventh Mental Measurement Yearbook, 1972, p. 736.

Hopkins[1] found "the WRAT and the Gilmore Accuracy Reading sub-test to convey almost identical and interchangeable information." This is reflected in Table #1, which shows no increase in the amount of variance for the Gilmore Accuracy test. Garlock, et al,[2] also found the relationship of both of the above tests with the Comprehension section of the Gilmore "substantial, yet low enough to clearly illustrate the need for both types of tests."

The Similarities sub-test of the WISC ranked next and accounted for .048 percent of the variance. The simple r was .350 and the r^2 increased to .653 when the Gilmore and WRAT were added. This sub-test measures verbal concept formation, logical thinking, ability to abstract and make inferences, and ability to categorize by seeing and relating essentials. Spache,[3] in his definition of reading, listed as one of the factors in comprehension that of "analysis-synthesis of ideas" which he also called, at various points, "inductive reasoning,"

---

[1]Hopkins, Education and Psychological Measurement, XXII, 1962, pp. 791-793.

[2]Garlock, California Journal of Educational Research, Vol. 16, 1965, pp. 54-57.

[3]Spache, Toward Better Reading, 1963, pp. 62-63.

or "deductive thinking and problem solving," or "abstract reasoning."

The Durrell Comprehension sub-test and the Digit Span sub-test of the WISC both contributed .01 percent to the variance of the PEP test.

The Gilmore Accuracy test, the Comprehension, Vocabulary, Information and Arithmetic sub-tests of the WISC, the Vocabulary of the Durrell, and the Cloze tests did not significantly increase the amount of variance accounted for in the PEP test inasmuch as all of the correlation that these tests have with the PEP test had already been accounted for by correlating them with the WRAT, Gilmore Comprehension tests, Similarities Sub-test, Durrell Comprehension and Digit Span Sub-test.

An analysis employing the same variables and adding sex was then computed. The simple r for Sex was -.100 (the r^2 .677) and the r^2Q was .001 percent, which accounts for no significant increase in variance. This concurs with the findings of Broom. Broom's[1] conclusion regarding sex and reading was that "sex differences were found to be not statistically significant."

---

[1]Broom, Journal of Educational Research, XXIV, 1941, p. 593.

When Verbal IQ was introduced in the third regression analysis, the simple r was .547. Adding Verbal IQ contributed .017 percent to the total variance, which was not significant.

In Davis' study,[1] reasoning and word knowledge emerged as the first and second factors.

This investigator would agree with Lennon,[2] who found a lack of consistency in his findings with respect to the generality or specificity of comprehension abilities. His studies

> . . . agree that most of the measurable variance in tests of reading competency . . . can be accounted for in terms of a fairly small number of factors, certainly not more than six being required to account for better than 90 percent of the variance.

Lennon also found that:

> Numerous superficially discrete reading skills to which separate names or titles have been attached are in fact so closely related, as far as any test reveals, that we must consider them virtually identical.

The WRAT test calls for the identification of words in isolation. The Gilmore Comprehension test calls for the ability to draw inferences from what is

---

[1]Davis, Psychometrika, 1944, IX, pp. 185-187.

[2]Lennon, Reading Teacher, 1962, Vol. 15, pp. 326-327.

read, to derive meaning of words from context, and to perceive main ideas. The Similarities sub-test of the WISC requires the ability to do abstract reasoning. These three tests were found to have the greatest similarity with the PEP test and accounted for 65 percent of the variance. The remaining tests, we can then assume, had tasks which were so closely related to these three tests and to each other that they added nothing further to account for the variance of the PEP test.

However, of the six reading tests administered, all evidenced significant correlations with the PEP test. This suggests a positive relationship among these measures.

Of the WISC correlations with the PEP test, five of them were found to have significant correlations, and this suggests it has concurrent validity. Digit Span correlation was the only one that was not significant.

This investigation has shown that the PEP test correlates with oral and silent reading tests and with tests of intellectual ability and has concurrent validity with these measures.

It has also demonstrated that the PEP test is measuring the three components of reading comprehension

ability claimed for it: comprehension of explicitly stated material, comprehension of implicit or latent material, and word knowledge.

Section Three

Results

This study was centered around fifteen null hypotheses which reflected the investigator's expectations with regard to the concurrent validity of the PEP test with individualized and standardized tests of reading and intelligence. The following results were obtained.

1. Hypothesis #1:- The correlation between the PEP test and the Vocabulary sub-test of the Durrell Reading test was ($r = .59$ $p < .01$), significant at the .01 level.

2. Hypothesis #2:- The correlation between the PEP test and the Comprehension sub-test of the Durrell Reading test was ($r = .55$ $p < .01$), significant at the .01 level.

3. Hypothesis #3:- The correlation between the PEP test and the total Durrell Reading test was ($r = .47$ $p < .01$), significant at the .01 level.

4. Hypothesis #4:- The correlation between the PEP test and the Accuracy sub-test of the Gilmore Oral Reading Test was ($r = .59$ $p < .01$), significant at the .01 level.

5. Hypothesis #5:- The correlation between the PEP test and the Comprehension sub-test of the Gilmore Oral Reading Test was ($r = .55$ $p < .01$), significant at the .01 level.

6. Hypothesis #6:- The correlation between the PEP test and the Reading sub-test of the WRAT was ($r = .71$ $p < .01$), significant at the .01 level.

7. Hypothesis #7:- The correlation between the PEP test and the Cloze test was ($r = .51$ $p < .01$), significant at the .01 level.

8. Hypothesis #8:- The correlation between the PEP test and the Information Scaled Score of the WISC was ($r = .44$ $p < .01$), significant at the .01 level.

9. Hypothesis #9:- The correlation between the PEP test and the Vocabulary Scaled Score of the WISC was ($r = .57$ $p < .01$), significant at the .01 level.

10. Hypothesis #10:- The correlation between the PEP test and the Arithmetic Scaled Score of the WISC was ($r = .32$ $p < .05$), significant at the .05 level.

11. Hypothesis #11:- The correlation between the PEP test and the Similarities Scaled Score of the WISC was ($r = .35$ $p < .02$), significant at the .02 level.

12. Hypothesis #12:- The correlation between the PEP test and the Comprehension Scaled Score of the WISC was ($r = .29$ $p < .05$), significant at the .05 level.

13. Hypothesis #13:- The correlation between the PEP test and the Digit Span Scaled Score of the WISC was ($r = .19$), which is not significant.

14. Hypothesis #14:- The correlation between the PEP test and the total Verbal IQ of the WISC was ($r = .54$ $p < .01$), significant at the .01 level.

15. Hypothesis #15:- The correlation between the PEP test and sex as a factor was ($r = .-10$), which is not significant.

CHAPTER V

SUMMARY, CONCLUSIONS AND IMPLICATIONS

This chapter contains the summary, conclusions and implications of the investigation. The summary includes the purpose and findings of the investigation. The conclusions and implications for future research drawn from the findings are also presented.

Summary

The main purpose of this study was to investigate the concurrent validity of the third grade PEP test with standardized tests. The tests chosen for this study were the Durrell Reading Test, the Gilmore Oral Reading Test, the WRAT, the WISC Verbal IQ, including the following verbal sub-tests: (Information, Vocabulary, Arithmetic, Similarities, Comprehension, and Digit Span) and the Cloze Test constructed by the investigator.

Several questions concerning the relationship of concurrent validity to other measures were posed and answered.

The relationship between the PEP test and several variables was established. The intellectual abilities that correlated highest were also established.

The three components of reading comprehension ability the PEP test was reputed to measure, (comprehension of explicitly stated material, comprehension of implicit or latent meaning, and word knowledge) were analyzed.

Several interesting findings were revealed relative to the standardized tests that accounted for the highest degree of variance with the PEP test.

The test that correlated the highest with the PEP test was the WRAT. The WRAT is a measure of word recognition, i.e., sight vocabulary. The relationship that this test has with other tests is possibly accounted for by the amount of variance resulting from the factor of word knowledge or some more global factor common to all of them.

The relationship between the PEP test and tests of paragraph reading-comprehension was also measured by administering standardized tests of both oral and silent paragraph reading. The test that, when added to the

multiple R, significantly increased the amount of variance accounted for in the dependent variable (i.e., the PEP test) was the Comprehension sub-test of the Gilmore Oral Reading Test. This test is a measure of oral comprehension and word recognition. The relationship that this test has with other tests is probably the variance resulting from word knowledge and recall.

The WISC was used to identify those intellectual abilities that correlated highest in a multiple regression. The sub-test of the WISC that, when added to the multiple R, increased the amount of variance accounted for in the dependent variable was the Similarities test. This test is a measure of abstract reasoning and making inferences. The relationship that this test has with other tests is probably the variance resulting from making inferences and abstract reasoning.

Sex differences were not found in any measure of the study. Thus, it is concluded that the variable of sex had little or no apparent effect on the measured results of the study.

Similarly, Verbal IQ did not add anything significant to the variance.

The administration of the PEP test yields results which could also be obtained by administering a battery of those tests with which it is correlated.

The results are a measure of comprehension and word knowledge in the same sense as would be yielded by the administration of such a battery of tests.

On the basis of this analysis, it would appear that the State of New York is justified in stating that the PEP test is in fact a measure of reading comprehension and word knowledge.

Implications for Future Research

According to Massey,[1] "it has been known, for many years, that a breakdown in reading ability seems to occur somewhere between the primary grades and fifth grades." More studies should be focused on the third grade since this appears to be a critical grade.[2]

This investigation should be replicated with different socio-economic groups in order to determine

---

[1] Charles Henry Massey, "A Comparison of Third Grade Reading Practices, Objectives, and Achievement Test Results Between A Conventional Program and A Behavioral Objective Approach Program," (unpublished, Ed.D. dissertation, Walden University).

[2] Ibid.

the extent to which ethnic and economic background have an affect on test scores.

A larger sample should be employed in order to determine whether it will yield different results.

Another study using other standardized tests with the same dependent variable should be undertaken to see whether there will be different results.

A factor analysis on the content validity of the PEP test should be undertaken in order to establish what the constructs are.

Conclusions

The administration of the PEP test yielded results which could also be obtained by administering a battery of those tests with which it is correlated.

The results are a measure of comprehension and word knowledge in the same sense as would be yielded by the administration of such a battery of tests.

On the basis of this analysis it would appear that the State of New York is justified in stating that the PEP reading test is measuring the three components of reading ability claimed for it: comprehension of explicitly stated material, comprehension of implicit or latent meaning, and word knowledge.

APPENDIX

Table 6

CENSUS

GENERAL SOCIAL AND ECONOMIC CHARACTERISTICS,
PLAINVIEW, LONG ISLAND, NEW YORK, 1970[1]

Total Population	33,188
English, Mother Tongue	22,631
Family Composition, Families	7,877
Income, Median	16,198
Mean	17,472
Income, Less Than Poverty	1.7%
Negro Population	Less than 400
Government Workers	
Male	1,131
Female	885
Professional, Technical & Kindred Workers	2,942
Managers & Administrators, except Farmers	1,754
Teachers, Elementary & Secondary	665
Total Employed 16 years old or over	12,616
College, 4 years or more	1,152
Median School Years Completed	12.5
High School Graduates	75.2%
Female, 25 years old or over	75.2%
Male, 25 years old or over	73.5%
School Enrollment	
Kindergarten	598
Kindergarten, Public	576
Elementary	7,084
Elementary, Public	6,135
Secondary	3,746
Secondary, Public	3,478
College	1,392

As can be seen from the data in Tables 1 and 2 the sample was drawn from a middle-class population.

[1] Census of Population, 1970, United States Department of Commerce, Bureau of the Census, Final Report PC (1) C34 New York, U. S. Government Printing Office, Washington, D.C., 1972.

FORMAT AND SAMPLE PAGES OF

TESTS USED IN THE

INVESTIGATION

Wide Range Achievement Test

Title	Wide Range Achievement Test[1]
Author:	J. F. Jastak and S. B. Jastak
Publisher:	Guidance Associates
Date:	December, 1965
Form:	Level 1 children 9 months to 11 years (Also includes Arithmetic, Spelling, Writing)
Format:	Reading, orally naming 2 letters in written or printed name, identifying 10 letters by form, naming 13 letters of the alphabet, and pronouncing 75 words.
Administration:	The examiner points to the words and asks the subject to look at them and say them aloud. Time is 10 seconds per word.
Reliability:	Split-half correlation coefficients and standard errors of measurement at age 7 is .993 and age 8 .991.
Validity:	New Stanford Paragraph Reading (1946) N= 389 + .81 New Stanford Word Reading (1946) (7th and 8th grade) + .84 Woody-Sangren Silent Reading Test Wagner's study 29 children + .78 McCoy's study 57 children + .74

[1] Jastak and Jastak, p. 39.

COPYRIGHT 1965 by ... Printed in U.S.A.

Percentiles and Standard Scores corresponding to grade rating and age may be found in the Manual.

Level I—Reading—Grade Norms

Score	Grade	Score	Grade	Score	Grade	Score	Grade	Score	Grade	Score	Grade	Score	Grade
1	N.3	16-17	Kg 6	36-37	1.9	53	3.3	66	5.3	79	8.1	92	12.9
2	N.8	18	Kg 7	38	2.0	54	3.5	67	5.5	80	8.4	93	13.3
3	Pk.1	19-20	Kg 8	39-40	2.1	55	3.6	68	5.7	81	8.7	94	13.7
4	Pk.2	21	Kg 9	41	2.2	56	3.8	69	5.9	82	9.0	95	14.1
5	Pk.4	22	Gr.1.0	42-43	2.3	57	3.9	70	6.1	83	9.3	96	14.5
6	Pk.5	23	1.1	44	2.4	58	4.1	71	6.3	84	9.7	97	14.9
7	Pk.7	24-25	1.2	45-46	2.5	59	4.2	72	6.5	85	10.1	98	15.4
8	Pk.9	26-27	1.3	47	2.6	60	4.4	73	6.7	86	10.5	99	15.8
9	Kg 1	28-29	1.4	48	2.7	61	4.5	74	6.8	87	10.9	100	16.2
10-11	Kg 2	30-31	1.5	49	2.8	62	4.7	75	7.0	88	11.3		
12	Kg 3	32-33	1.6	50	2.9	63	4.8	76	7.2	89	11.7		
13-14	Kg 4	34	1.7	51	3.0	64	5.0	77	7.5	90	12.1		
15	Kg 5	35	1.8	52	3.1	65	5.1	78	7.8	91	12.5		

Level II—Reading—Grade Norms

Score	Grade	Score	Grade	Score	Grade	Score	Grade	Score	Grade	Score	Grade	Score	Grade
0	Pk.3	16	1.3	29	4.4	44	6.8	55	9.3	6	13.0	81	16.8
1	Pk.8	17	1.5	30	4.6	45	6.9	56	9.6	6	13.4	82	17.1
2	Kg 1	18	1.7	31	4.8	46	7.1	57	9.9	7	13.5	83	17.4
3-4	Kg 2	19	1.8	32	5.0	47	7.3	58	10.2	7	13.8	84	17.7
5-6	Kg 3	20	2.0	33	5.2	48	7.5	59	10.5	7	14.1	85	18.0
7	Kg 4	21	2.2	34	5.4	47	7.7	60	10.8	7	14.4	86	18.3
8	Kg 5	22	2.4	35	5.6	48	7.9	61	11.3	7	14.7	87	18.6
9	Kg 6	23	2.6	36	5.8	49	8.1	62	11.6	7	15.0	88	19.0
10-11	Kg 7	24	2.8	37	6.0	50	8.3	63	11.9	7	15.3	89	19.3
12	Kg 8	25	3.2	38	6.2	51	8.5	64	12.2	7	15.6		
13	Kg 9	26	3.5	39	6.3	52	8.7	65	12.5	7	15.9		
14	Gr.1.0	27	3.9	40	6.5	53	8.9	66	12.8	7	16.2		
15	1.1	28	4.2	41	6.6	54	9.1	67	12.8	8	16.5		

LEVEL 2

Two letters in name (2) A B O S E R T H P I U Z Q (13) 13

milk city in tree animal himself between chin split form 23

misogyny beneficent desuetude egregious heinous internecine synecdoche 89

LEVEL 1

cat see red to big work book eat was him how 36

seismograph spurious idiosyncrasy itinerary pseudonym aborigines 100

A R Z H I Q S E O

Two letters in name (2) A B O S E R T H P I U Z Q

Durrell Listening-Reading Series

Title:	Durrell Listening-Reading Series[1] Test Three--Vocabulary Reading Test Four--Paragraph Reading
Author:	Donald D. Durrell
Publisher:	Harcourt, Brace, Jovanovich, Inc.
Date:	1970
Form:	DE for grades 3-5-6
Format:	Vocabulary Test--The child has to relate a word to one of four categories in a cluster which are illustrated by pictures. There are eight parts of 12 words each. After reading a paragraph silently, the child has to relate a word to one of four categories in a cluster. The category words constituting the opinions in each cluster are illustrated by pictures in order to remind the child of their meaning. There are eight parts of eight sentences each.
Administration Time:	Vocabulary 40 minutes Comprehension 45 minutes
Reliability Data:	Kuder-Richardson formula 21 for grade 3 is vocabulary .84 and paragraph reading .81. Split-half, odd-even reliability corrected is vocabulary .89 and paragraph reading .85.

---

[1] Donald D. Durrell, Manual for Listening and Reading Tests (New York: Harcourt Brace, Jovanovich, Inc. 1970), p. 15.

Validity:

Correlations at 3rd grade with Word Knowledge and Reading sub-tests of Metropolitan Achievement Test.

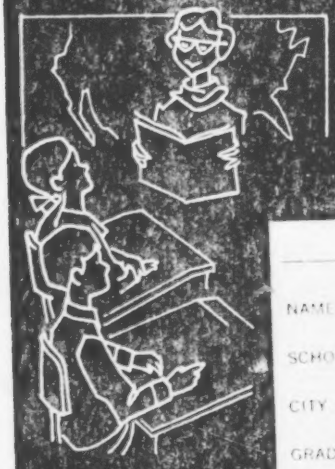
N 246	Vocabulary Reading	Paragraph Reading
MAT Word Knowledge	.64	.56
MAT Reading	.57	.58

Durrell Listening - Reading Series

DONALD D. DURRELL · MARY B. BRASSARD

INTERMEDIATE LEVEL — Form DE

COMBINED LISTENING AND READING TESTS – Hand-Scorable Edition



PUPIL IDENTIFICATION BOX

NAME _____ last _____ first _____ initial

SCHOOL _____

CITY _____ STATE _____

GRADE _____ TEACHER _____

DATE OF BIRTH _____ yr. mo. day BOY ☐ GIRL ☐ AGE _____ yrs. mos.

SCORE BOX

Listening Test						Reading Test					
Vocabulary		Paragraphs		Total		Vocabulary		Paragraphs		Total	
1	2	3	4	5	6	7	8	9	10	11	12
Number Right	Potential Reading Grade Equiv.	Number Right	Potential Reading Grade Equiv.	Number Right	Potential Reading Grade Equiv.	Number Right	Actual Reading Grade Equiv.	Number Right	Actual Reading Grade Equiv.	Number Right	Actual Reading Grade Equiv.

Date of Testing _____

Grade Placement _____ yr. _____ mo.

Date of Testing _____

HARCOURT BRACE JOVANOVICH, INC. NEW YORK

Copyright © 1970 by Harcourt Brace Jovanovich, Inc. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher. Printed in U.S.A.

Reading Test

TEST 4: PARAGRAPH READING

DIRECTIONS

Here are some stories which tell about two people, places, or things which are alike in some ways but different in other ways. After each story there are statements for you to classify. You do this by filling in the answer space under the right heading. The sample will be done for you.

SAMPLE

Bill and George are two boys who are neighbors and go to the same school. Bill likes sports and is good in all of them. George doesn't care for sports; he likes to play the piano and listen to music.

Bill's favorite sport is football. He is quarterback on the school team. He would like to be a professional football player some day.

Although George likes the piano best, he plays clarinet in the school band. When the band plays at football games, the boys ride home together after the game.

	True only of Bill	True only of George	True of both	Answer is not given
S1 He plays football.	S1 ●	○	○	○
S2 He goes to school.	S2 ○	○	●	○
S3 He earns money after school.	S3 ○	○	○	○
S4 He plays the clarinet.	S4 ○	○	○	○

PUPIL INFORMATION BOX

NAME _____
last first initial

BOY ☐ GIRL ☐ GRADE _____ TEACHER _____

SCHOOL _____

CITY _____ DATE OF BIRTH _____
yr. mo. day

STATE _____ AGE _____
yrs. mos

SCORE BOX

Listening Test						Reading Test					
Vocabulary		Sentences		Total		Vocabulary		Sentence		Total	
1 Number Right	2 Potential Reading Grade Equiv.	3 Number Right	4 Potential Reading Grade Equiv.	5 Number Right	6 Potential Reading Grade Equiv.	7 Number Right	8 Actual Reading Grade Equiv.	9 Number Right	10 Actual Reading Grade Equiv.	11 Number Right	

Date of Testing _____

Grade Placement _____
yr. mo.

Date of Testing _____

New York State Pupil Evaluation Program
in Reading Grade 3

Title:	New York State Test in Reading[1] Beginning Grade 3 Pupil Evaluation Program
Author:	New York State Education Department Bureau of Pupil Testing and Advisory Services
Date:	October, 1973
Form:	C
Format:	50 Multiple Choice questions of 4 choices each. Twelve passages of graded paragraphs according to difficulty to test comprehension of explicitly stated material, of implicit or latent meaning, word knowledge and pronoun antecedent
Administration Time:	45 minutes
Reliability Data:	This is a new test and none was furnished by the State Department. The investigator computed one by tallying number of each score of odd answers correct correlated and even scores correct correlated using Spearman-Brown correction formula. The reliability was .91.

[1] School Administrator's Manual, The State Education Department Bureau of Pupil Testing and Advisory Services (Albany, New York, Fall, 1973).

Validity:

In October, 1973, these new tests were administered to all pupils. Representative samples of pupils in public and nonpublic schools also were administered the Form A editions, which were previously used in the PEP program. The grade 3 sample consisted of 7,276 pupils. The norms were prepared by equating the performance of these pupils on the new tests with their performance on the old tests. The norms provided an estimate of the distribution of scores that would have been obtained if the new tests had been administered in the fall of 1966.

New York State Test
in
READING

Beginning Grade **3**

Form **C**

Name..... Date of Birth.....

School..... Date.....

City.....

Teacher.....

Raw Score	Percentile Rank

The University of the State of New York
The State Education Department
Albany, New York 12224

New York State Test in READING

Beginning Grade 3 Form C

This is a test to see how well you understand what you read. It contains 13 stories. You are to read each story and answer the questions that follow it. The right answer for each question will be found among the choices that follow the question. After you find the right answer make an $\times$ in the box in front of it.

To make sure that you know what to do, read the sample story below.

SAMPLE STORY

The boys and girls were happy when summer came. It was a time for swimming.

A In this story, the boys and girls were

- ☐ tired
- ☒ happy
- ☐ sad
- ☐ angry

B The word it means

- ☐ summer
- ☐ swimming
- ☐ time
- ☐ playing

You will have 45 minutes to finish the test. Try to answer as many questions as you can even if you are not sure that your answers are right.

DO NOT TURN THE PAGE.

Copyright 1973
The University of the State of New York
The State Education Department
Albany, New York 12224

School was over for the day. Pat was on her way home. She was alone. She was a new girl and had not yet made any friends.

1 This story is mainly about

- ☐ home
- ☐ school
- ☐ a new girl
- ☐ Pat's friends

2 Where was Pat going?

- ☐ to her home
- ☐ to school
- ☐ to the park
- ☐ to see a friend

3 This story most likely took place in the

- ☐ morning
- ☐ afternoon
- ☐ evening
- ☐ night

Carlos played by himself one rainy day. He made a train. His train was very long. It was made of cartons, boxes, pots, and pans. His kitten rode in his train.

4 When Carlos made his train, one thing he used was

- ☐ chairs
- ☐ cans
- ☐ blocks
- ☐ pans

5 This story is mainly about

- ☐ who rode in Carlos' train
- ☐ what Carlos did on a rainy day
- ☐ where Carlos went on a rainy day
- ☐ how Carlos rode on a train

☞ GO RIGHT ON TO THE NEXT PAGE.

Gilmore Oral Reading Test

Title:	Gilmore Oral Reading Test[1]
Author:	John V. Gilmore and Eunice C. Gilmore
Publisher:	Harcourt, Brace & World, Inc.
Date:	1968
Form:	D
Format:	Ten paragraphs ascending from first grade level, at a grade per paragraph. Each tells a continuing story about a family, with each paragraph a little longer, more mature, and more difficult than the preceding one. The Accuracy score is based on fluency as well as of accuracy of word recognition. The Comprehension score is based on the answers to five questions per paragraph.
Administration:	Subject is timed with a stopwatch.
Reliability:	Accuracy--5 children were tested. The accuracy in third grade was .94. The comprehension score for third grade was .60.
Validity:	This test score was compared with those obtained from other oral reading tests. To obtain such data, for the original forms, the Gilmore Oral Reading Test, the Standardized Oral Reading Paragraphs by Gray, and the oral reading test from the Durrell Analysis of Reading Difficulty were

[1] John V. Gilmore and Eunice C. Gilmore, Manual of Directions for Gilmore Oral Reading Test, (New York: Harcourt, Brace & World, Inc., 1968).

administered to 24 pupils of the same age in grade 5. The Gilmore-Gray had a .77 validity, the Gilmore-Durrell a .80 correlation and .59 comprehension and the Gray-Durrell a .73 correlation.

RECORD BLANK

Form D

Gilmore Oral Reading Test

by JOHN V. GILMORE
EUNICE C. GILMORE

Name _____ Age _____

Address _____

School _____ Grade _____

Date of Testing _____

Examiner _____

HARCOURT, BRACE & WORLD, INC.
NEW YORK CHICAGO SAN FRANCISCO ATLANTA DALLAS

Copyright © 1968 by Harcourt, Brace & World, Inc. All rights reserved. Printed in U.S.A.

FORM D-1

I can see Bob.
Spot is his dog.
Bob has a sister.
She is Jane.
Jane has a cat.
Bob and Jane like pets.

TIME_____Seconds

1. What is the boy's name? +
2. Who is Spot? +
3. What is the name of Bob's sister? +
4. What pet does she have? —

___ 5. How do Jane and Bob feel about pets? —

ERROR RECORD	Number
Substitutions	
Mispronunciations	
Words pronounced by examiner	
Disregard of punctuation	
Insertions	
Hesitations	
Repetitions	
Omissions	
Total Errors	

NAME_____

TEST SUMMARY

Form D

PARAGRAPH	ACCURACY		COMPREHENSION	RATE	
	ERRORS	10 MINUS NO. ERRORS	NO. RIGHT (OR CREDITED)	WORDS IN ¶	TIME IN SEC.
1				24	
2				45	
3				50	
4				71	
5				97	
6				120	
7				126	
8				157	
9				194	
10				248	
	ACC. SCORE (TOT. "10 MINUS NO. ERRORS" COLUMN)		COMP. SCORE (TOT. NO. RIGHT OR CREDITED)	(1) NO. WORDS READ*	
				(2) TIME IN SEC.*	
STANINE				(1) ÷ (2)	X 60
GRADE EQUIV.				RATE SCORE (WPM)	
RATING					

*Do not count "reading" paragraph or paragraphs below "basal."

COMMENTS:

Cloze Test

The selection of a passage from a basal reader[1] was done by the investigator. The first sentence was left intact. Following that, a word was deleted after every fourth word, and a blank space of fifteen spaces was substituted. The last sentence was also left intact. There were fifty-two deletions. The test was reproduced by mimeographing and then collating the two pages. The tests were administered by the investigator to each of the third grade classes. Each subject recorded his or her response in the space left blank by the deletions. The score was obtained by counting the number of correct responses. This was the raw score.

---

[1] William D. Sheldon and Mary D. Austin, Magic Windows (Allyn Bacon: 1959), pp. 97-100.

Cloze Test

Name_____Teacher_____Raw Score____

THE CIRCUS BAND LEADER

Some boys and girls believe that clowns are the most important people in the circus. Others say the acrobats_____the cowboys are the_____ important. Sometimes the circus_____leader is forgotten, but _____is really a very_____ man.

The band leader must_____when to play fast_____when to play slowly. _____ must know when to_____loudly and when to _____softly. He must know_____kind of music to_____when horses trot. _____ knows just when_____clown will jump on_____ horse's back. He knows, _____, what kind of music _____needs for the acrobats. _____ conductor tells the whole_____how to work and _____to work.

Mr. Tiwarski _____, band leader for the _____Bill Brothers' Circus. The _____ people called him Mr._____because it took too _____to say Mr. Tiwarski. _____people are busy people. _____don't like to take _____time to say long_____like Tiwarski.

Even the_____knew that he was _____good leader. Mr. T. _____just when to point_____the drums to play_____ elephant music or when_____point to the flutes _____the monkey music.

All_____circus people and the_____ ____ animals liked Mr. T., _____. He smiled most of _____time and his gay_____made everyone else smile.

_____was surprised on the_____ Mr. T. appeared without_____smile. People were

even _____surprised when he pointed_____ __

the flutes to play_____elephant music. Then the

_____played the monkey music.

_____elephants appeared to be_____ ____,

too. The boys and_____and the mothers and

_____did not know what_____happened.

Even the men who played the flutes were surprised.

Wechsler Intelligence Scale for Children

Title:	Wechsler Intelligence Scale for Children[1]
Author:	David Wechsler
Publisher:	The Psychological Corporation
Date:	1949
Format:	Ages 5-15--Verbal (Information, Comprehension, Arithmetic, Similarities, Vocabulary, Digit Span). There are 30 Information questions, 14 Comprehension, 16 Arithmetic, 16 Similarities, and 9 Digits forward and 9 Digits backward and 40 Vocabulary Meanings. The subject has to make either 3 or 5 errors in the various categories before testing is stopped.
Administration:	Subject is timed with a stopwatch.
Reliability:	Reliability coefficients have been computed by the split-half technique with appropriate correction. For Information .66, Comprehension .59, Arithmetic .63, Similarities .66, Vocabulary .77, Digit Span .60, Verbal Score without Digit Span .88. For Digit Span the Spearman-Brown formula was used.

David Wechsler, Manual of Directions, Wechsler Intelligence Scale for Children, Psychological Corporation.

WISC RECORD FORM

NAME _____ AGE ____ SEX ____

ADDRESS _____

PARENT'S NAME _____

SCHOOL _____ GRADE _____

REFERRED BY _____

	Year	Month	Day
Date Tested			
Date of Birth			
Age			

	Scaled Score	IQ
Verbal Scale	*	
Performance Scale	*	
Full Scale		

*Prorated if necessary

	Raw Score	Scaled Score
VERBAL TESTS		
Information		
Comprehension		
Arithmetic		
Similarities		
Vocabulary		
(Digit Span)		
Sum of Verbal Tests		
PERFORMANCE TESTS		
Picture Completion		
Picture Arrangement		
Block Design		
Object Assembly		
Coding		
(Mazes)		
Sum of Performance Tests		

NOTES

_____
Examiner

Copyright 1949, All rights reserved.
The Psychological Corporation
522 Fifth Avenue
New York 36, N. Y.

Printed in U. S. A. 0130

57-143 AS

REFERENCES

Books

Buros, Oscar Krisen (ed.). The Seventh Mental Measurement Yearbook. Highland Falls, New Jersey: The Gryphon Press.

Dechant, E. V. Improving the Teaching of Reading. Englewood Cliffs, New Jersey: Prentice Hall, 1964.

Directions for Administering and Scoring. New York State Pupil Evaluation Program "PEP" Tests in Reading Beginning Grades 3 and 6, New York State Education Department, Albany, 1973.

Durrell, Donald D. and Brassard, Mary B. Manual for Listening and Reading Tests. Durrell Listening-Reading Series, Intermediate Level. New York: Harcourt, Brace, Jovanovich, Inc., 1970.

Gilmore, John V. and Eunice C. Gilmore. Manual of Directions. Gilmore Oral Reading Test. New York: Harcourt, Brace & World, Inc., 1968.

Guilford, J. P. Psychometric Methods. New York: McGraw Hill Book Company, 1934.

Inhelder, B. and Piaget, S. The Growth of Logical Thinking From Childhood to Adolescence. London: Routledge and Kegan, 1958.

Jastak, J. F. and Jastak, S. R. Manual of Instructions. The Wide Range Achievement Test. Wilmington, Delaware: Guidance Associates, 1965.

Kerlinger, Fred M. and Pedhazer, Eliza J. Multiple Regression and Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1973.

Kraus, Philip E. A Longitudinal Study of Children from Kindergarten into the Adult Years. New York: John Wiley and Sons. Cited in New York Teacher, April 1, 1973.

Lindvall, C. M. Measuring Pupil Achievement and Aptitude. New York: Harcourt, Brace & World, 1966.

Robinson, Helen M. Why Pupils Fail in Reading. Chicago, University of Chicago Press, 1966.

School Administrator's Manual. The State Education Department, Bureau of Pupil Testing and Advisory Services. Albany, New York, 1973.

Sheldon, William D. and Austin, Mary D. Magic Windows. Allyn Bacon, 1959.

Spache, George D. Toward Better Reading. Champaign, Illinois: Garrard Publishing Company, 1963.

Smith, H. P. and Dechant, E. V. Psychology in Teaching Reading. Englewood Cliffs, New Jersey: Prentice Hall, 1961.

Thayer, B. T. Public Education and Its Critics. New York: MacMillan Company, 1969.

Wechsler, David. Manual of Directions. Wechsler Intelligence Scale for Children. New York: Psychological Corporation.

Articles

Bormouth, John R. "Durrell Listening Reading Series," in The Seventh Mental Measurement Yearbook, Oscar Krisen Buros (ed.), Highland Falls, New Jersey: The Gryphon Press, 1972.

Broom, M. E. "A Study of the Race and Sex Differences in Reading Comprehension," Journal of Educational Research, XXXIV (April, 1941), p. 593.

Conklin, R. C. and Dockrell, W. B. "The Predictive Validity and Stability of W.I.S.C. Scores over a Four Year Period," Psychology in the Schools, Vol. 4 (July, 1967), pp. 263-266.

Garlock, Jerry, Dollarhide, Robert S. and Hopkins, Kenneth D. "Comparability of Scores on the Wide Range and the Gilmore Oral Reading Tests," California Journal of Educational Research, Vol. 16 (March, 1965), pp. 54-57.

Hall, W. E. and Robinson, F. P. "An Analytical Approach to the Study of Reading Skills," Journal of Educational Psychology, XXXVI (1945), pp. 429-442.

Harris, J. Albert. "The Gilmore Oral Reading Test," in The Seventh Mental Measurement Yearbook, Oscar Krisen Buros (ed.), Highland Falls, New Jersey: The Gryphon Press, 1972.

Hopkins, Kenneth D., Dobson, James C. and Oldbridge, O.A. "The Concurrent and Congruent Validities of the Wide Range Achievement Test," Educational and Psychological Measurement, Vol. XXII, No. 4, 1962, pp. 791-793.

Hunt, C. M. Cited in "What Can Be Measured," Reading Teacher, Vol. 15, No. 5 (March, 1962).

Merwin, Jack. "The Wide Range Achievement Test," in The Seventh Mental Measurement Yearbook, Oscar Krisen Buros (ed.), Highland Falls, New Jersey: The Gryphon Press, 1972.

Naiven, Frederic B. and Bierbryer, Bruce. "Predicting Elementary School Children's Classroom Comprehension from Their WISC Results," Journal of Clinical Psychology, Vol. XXV, No. 1 (January, 1969), pp. 75-76.

Simpson, Robert L. "Reading Tests versus Intelligence as Predictors of High School Graduation," Psychology in the Schools, Vol. 7 (1970), pp. 363-365.

Smith, Kenneth B. "The Gilmore Oral Reading Test," in The Seventh Mental Measurement Yearbook, Oscar Krisen Buros (ed.), Highland Falls, New Jersey: The Gryphon Press, 1972.

Spache, George D. "Durrell Listening Reading Series," in The Seventh Mental Measurement Yearbook, Oscar Krisen Buros (ed.), Highland Falls, New Jersey: The Gryphon Press, 1972.

Stroud, J. B., Bloomers, P. and Lauber, Margaret. "Correlational Analysis of WISC and Achievement Tests," Journal of Educational Psychology, (1957,48), pp. 18-26.

Sundberg, Norman D. "The Practice of Psychological Testing in Clinical Services in the United States," American Psychologist, Vol. 16 (1960), pp. 79-83.

Taylor, Wilson L. "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, XXX (Winter, 1953), pp. 415-433.

_____. "Cloze Readability Scores as Indices of Individual Differences in Comprehension and Aptitude," Journal of Applied Psychology, XLI (February, 1957), pp. 25-26.

Thorndike, Robert T. "The Wide Range Achievement Test," in The Seventh Mental Measurement Yearbook, Oscar Krisen Buros (ed.), Highland Falls, New Jersey: The Gryphon Press, 1972.

Weaver, Wendell W. and Kingston, Albert J. "A Factor Analysis to the Cloze Procedure and Other Measures of Reading and Language Ability," The Journal of Communication, XII (December, 1963), pp. 252-261.

Wilson, Fran I. "Sex Differences in Beginning Reading in a Progressive School," Journal of Educational Research, XXXII (April, 1959), p. 570.

Newspapers

Newsday. June 21, 1974.

New York Post. January 21, 1974.

New York Teacher. April 1, 1973.

New York Times. May 27, 1973.

Unpublished Material

Bruininks, R. H. and Bruininks, in press (1969).

Davis, Frederick B. "Fundamental Factors of Comprehension in Reading" (Unpublished Ed.D. Dissertation, Howard University, 1941).

Eisner, Joseph. "The Plainview-Old Bethpage Community--A Survey" (Unpublished, Plainview-Old Bethpage Library, Plainview, 1968, mimeographed).

Horton, Raymond J. "The Construct Validity of Cloze Procedure" (Unpublished Ph.D. Dissertation, Hofstra University, 1973).

Massey, Charles Henry. "A Comparison of Third Grade Reading Practices, Objectives, and Achievement Test Results Between A Conventional Program And A Behavioral Objective Approach Program" (Unpublished Ed.D. Dissertation, Walden University).

Neville, D. and Bruininks, R. H. Appleton Century Crofts (New York: H. C. Haywood) citing Psychometric Intelligence, in press.

Wilson, G. M. cited in D. Neville, and R. H. Bruininks, in press.

Miscellaneous

Bulletin--Directions for Administering and Scoring New York State Tests in Reading Beginning Grades 3 and 6, Form C. (Albany, New York, Copyright, 1973).

Negotiating Bulletin Update. Plainview-Old Bethpage Federation of Teachers. (1972) (mimeographed).

Pupil Evaluation Program. New York State Education Department, Bureau of Pupil Testing and Advisory Services, February, 1974, letter.

Telephone Conversation with Jack A. Maybe, Chief, Bureau of Statistics, and Mr. Lott, Chief, Bureau of Testing, The University of the State of New York, The State Education Department, Albany, New York, 12224.

Traxler, Arthur E. "A Study of the Van Wagenen Dvorak Diagnostic Examination of Silent Reading Abilities," Educational Records Bureau Bulletin, No. 31 (January, 1941), p. 33.